

ARC Meeting
ALA Midwinter
January 7, 2011

WorldCat Quality

OCLC AMERICAS REGIONAL COUNCIL

Karen Calhoun

Vice President, Metadata Applications

Introduction by:

Betty Landesman, ARC Executive Committee
National Institutes of Health (NIH) Library



OCLC[™]

The world's libraries.
Connected.

GLOBAL COUNCIL

Beauty, like supreme
dominion,
is but supported
by opinion.

— Benjamin Franklin,
Poor Richard's Almanac

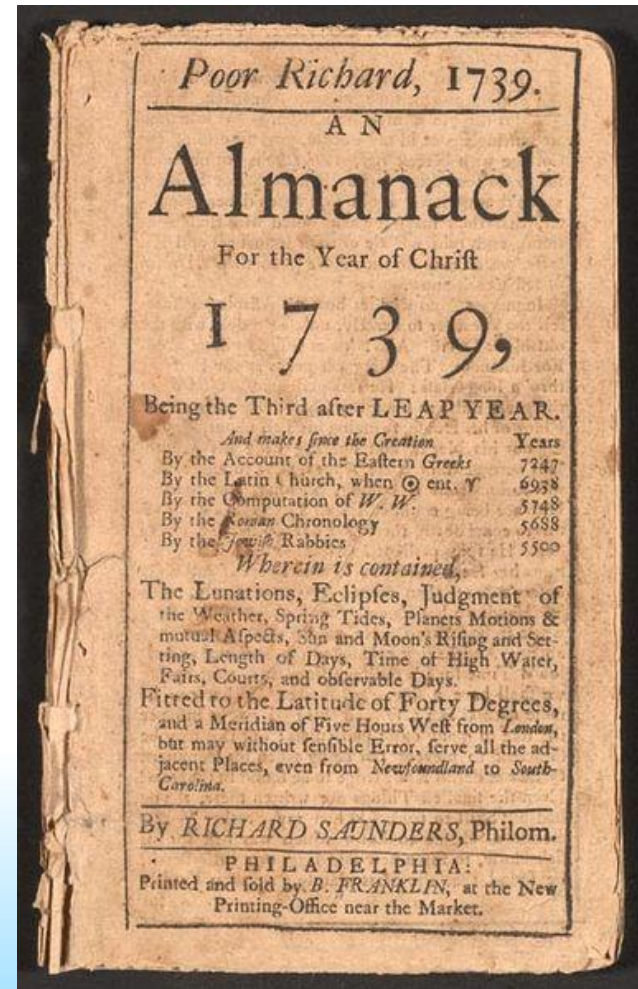


Image: public domain

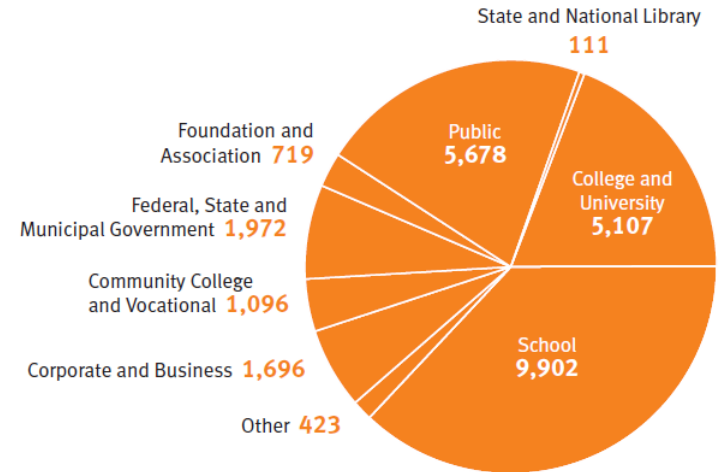
WorldCat Quality: Who is the audience?

- Library Professionals
 - Catalogers
 - Librarians of all types
- General Web Users
 - Expert searchers
 - End users (I just want to get stuff)
- Individuals doing known item searching

Who uses WorldCat?

- Libraries:

- 389.3 million items cataloged
- 57.8 million records added to WorldCat
- 10.2 million interlibrary loans arranged
- 68.4 million cataloging records exported



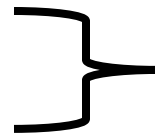
**Public, college/university,
State and national =
40%**

Source: OCLC annual report 2009/2010.

<http://www.oclc.org/news/publications/annualreports/2010/2010.pdf>

Who uses WorldCat.org?

Total End Users	7,583
Students*	28%
Teacher or professor	22%
Business professional	19%
Other	16%
Retired	9%
Healthcare professional	6%



69%
Students
Teacher/professor
Business
professional

*Type of Student	
Graduate college/Post-graduate	54%
Undergraduate college/Post-secondary	30%
High/Secondary school	10%
Middle/Intermediate/Junior high school	3%
Other	2%
Elementary/Primary school	1%

Source: *Online Catalogs: What Users and Librarians Want*, OCLC, 2009 (End-user pop-up survey)

Source: *Online Catalogs* study, PDF p. 16

<http://www.oclc.org/us/en/reports/onlinecatalogs/default.htm>

WorldCat.org traffic

2009/2010 annual report:

- 150 million click-throughs from partner sites to WorldCat.org
- 8.4 million click-throughs from WorldCat.org to library services

Objectives of our metadata quality research

- Start over without assumptions about what “quality” is
- Identify and compare metadata expectations - end user and librarian
- Define a new WorldCat quality program ...
- Taking into account the perspectives of all constituencies of WorldCat - end users and librarians

Online Catalogs: What Users and Librarians Want

End-Users expect online catalogs:

- to look/behave like popular Web sites
- to have summaries, abstracts, tables of contents
- to link directly to needed information

Librarians expect online catalogs:

- to help staff carry out work responsibilities
- to have accurate, structured data
- to exhibit library principles of organization

<http://www.oclc.org/us/en/reports/onlinecatalogs/default.htm>

Online Catalogs: What Users and Librarians Want



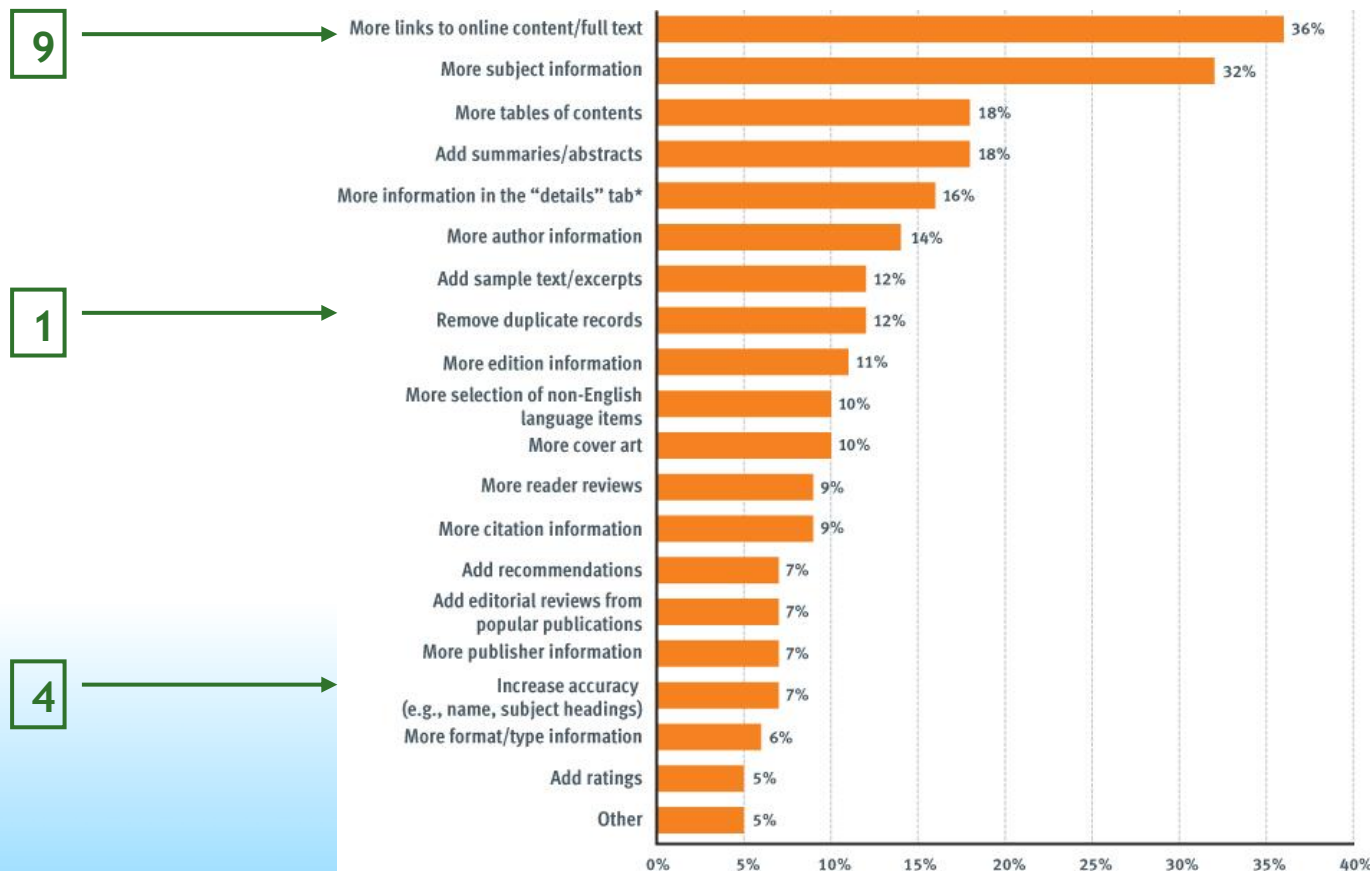
An OCLC Report



April 2009

Librarian/Staff Results: Highlighted Differences

End-User Results: Recommended Enhancements



Source: *Online Catalogs study*, PDF p. 51

Recommendations from librarian survey

- Merge duplicate bibliographic records
- Enrichment—TOCs, summaries, cover art—work with content suppliers, use APIs, etc.
- Make it easier to make corrections to records (fix typos; do upgrades); “social cataloging” experiment—Wikipedia
- More emphasis on accuracy/currency of library holdings

Composite view of what end users and librarians want—and what we are doing about it

Relative Ranking of Data Quality Enhancements in Library and End-user Surveys

Which of the following enhancements would you recommend? (Library survey)
*What changes would be **most** helpful to you in identifying the item that you need? (End-user pop-up survey)*

Comparable Enhancements Choice*	Relative Ranking		
	● Library Survey Respondents	● End-user Survey Respondents	
	Top Third	Middle Third	Bottom Third
Merge duplicate records	● #1 Choice	●	
More links to online content	● #1 Choice		
Add tables of contents to records	● ●		
Add summaries to records	● ●		
Add cover art to results	●	●	
Add more formats		●	●
More records for non-English materials		●	●
Add excerpts to the records	●		●

**Basis of 2009-2010
WorldCat Quality Program**

Source: *Online Catalogs study*, PDF p. 52

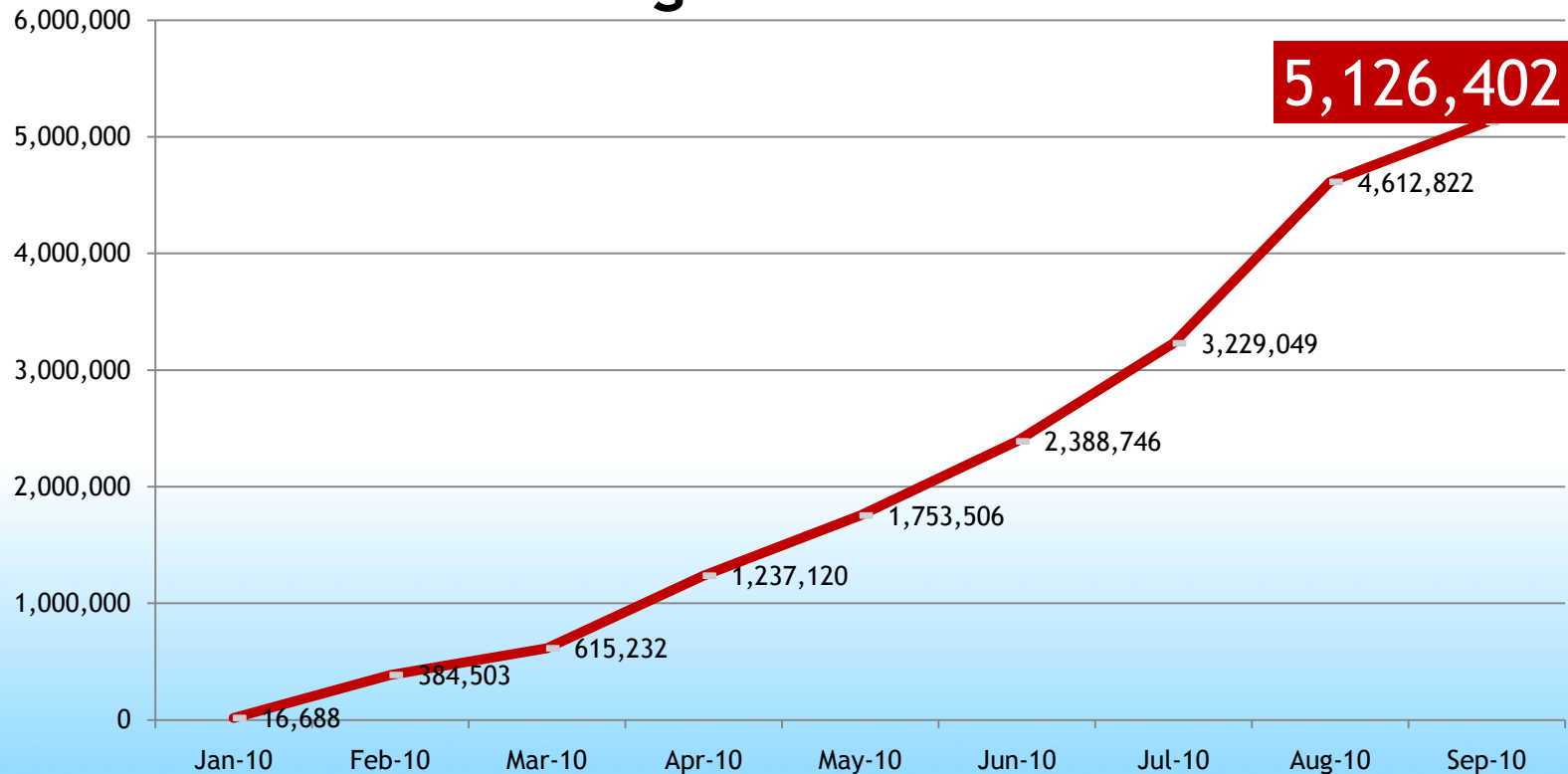
MERGE DUPLICATE BIBLIOGRAPHIC RECORDS

Duplicate Detection and Resolution (DDR) of WorldCat bibliographic records

- Reimplementation and expansion of previous software
 - Now handles all types of material (not just books)
- Fully operational in early 2010 in 2 separate processes
 - “Walking the database” (Complete September 2010)
 - Selected records from each day’s daily journal files (Ongoing)
- The result is “continuous cleaning” of WorldCat

Cumulative Number of Duplicates Removed, Jan.-Sept. 2010: 5.1 million

“Walking the Database”



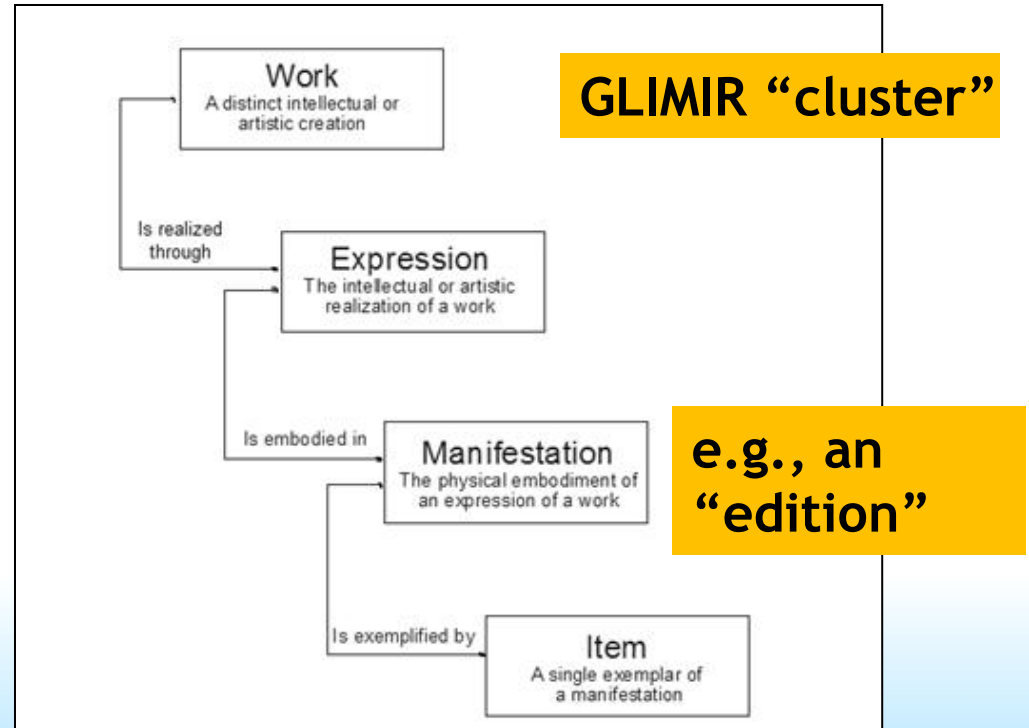
GLIMIR

GLIMIR = **G**lobal **L**ibrary **M**anifestation **I**dentifie**R**.

- Clusters manifestations and assigns unique identifier to each manifestation
- Clusters records for parallel records (differing languages of cataloging for the same manifestation) and for reproductions
- Re-clusters FRBR work sets

What is FRBR? What is a FRBR Work Set?

FRBR (Functional Requirements for Bibliographic Records) is a 1998 recommendation of the International Federation of Library Associations and Institutions (IFLA) to restructure catalog databases



Source: Figure 3.1, *Functional Requirements for Bibliographic Records: Final Report, 1998 text*. <http://www.ifla.org/VII/s13/frbr/frbr.htm>

GLIMIR

- Example: **Lieselotte Schwarz : Malerbücher**
 - Presently six records for same publication
 - In a GLIMIR version of WorldCat, two results instead of six
 - GLIMIR will convert six FRBR work sets to two

Search results for "Lieselotte Schwarz : Malerbücher"

Narrow by Format

- ✓ All Formats (6)
- ✓ Book (6)

Refine Your Search

Author

[Alexander Hildebrand](#) (6)
[Lieselotte Schwarz](#) (1)

Year

[1982](#) (6)
[1981](#) (1)

Language







[German](#) (4)
[Undetermined](#) (2)

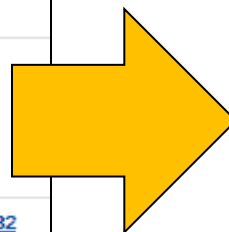
Results 1-6 of about 6 (.60 seconds)

[Select All](#) [Clear All](#)

Save to:



-  [Lieselotte Schwarz : Malerbücher](#)
by Alexander Hildebrand; Lieselotte Schwarz; Gutenberg-Museum (Mainz);
Book
Language: German
Publisher: Mainz : Gutenberg-Museum, 1982.
-  [Lieselotte Schwarz : Malerbücher](#)
by Alexander Hildebrand; Lieselotte Schwarz; Gutenberg-Museum Mainz.
Book
Language: German
Publisher: Mainz : Gutenberg-Museum, 1982, ©1981.
-  [Lieselotte Schwarz - Malerbücher](#)
by Alexander Hildebrand; Lieselotte Schwarz;
Book
Language: German
Publisher: Mainz : Gutenberg-Museum, 1982.
-  [Lieselotte Schwarz - Malerbücher : \[Ausstellung\] Gutenberg-Museum Mainz 1982](#)
by Alexander Hildebrand; Lieselotte Schwarz; Gutenberg-Museum. <Mainz>;
Book : Conference publication
Publisher: Mainz : Gutenberg-Museum, 1982.
-  [Lieselotte Schwarz, Malerbücher.](#)
by Alexander Hildebrand; Lieselotte Schwarz
Book
Publisher: Mainz : Gutenberg-Museum ; Wiesbaden : Schwarz, 1981.
-  [Malerbücher](#)
by Lieselotte Schwarz; Alexander Hildebrand
Book
Language: German
Publisher: Mainz : Gutenberg-Museum, 1982.



**GLIMIR
WILL
CLUSTER
THESE
FROM SIX
TO TWO
RESULTS**



OCLC™ The world's libraries.
Connected.
GLOBAL COUNCIL

WorldCat.org & WorldCat Local

Potential Usage of GLIMIR

- Non-English browser setting: If there is a matching manifestation cataloged in that language WorldCat.org & WorldCat Local can use that language record in place of the English language-of-cataloging record.
- Reproductions & originals represented on individual detailed records (e.g. display links to microform reproduction on the original manifestation.)
- GLIMIR clusters will greatly improve the accuracy of FRBR worksets, reducing appearances of exact duplicate records

MORE (BETTER) LINKS

Links to content: Load metadata for 4,500,000 eBooks from mass digitization, aggregator and publisher partners into WorldCat

	Year to Date (December)	Percent of Goal
Actual	4,326,436	96%

YTD Signees

**Project Gutenberg and Sage
BioDiversity Heritage Library loaded,
MyiLibrary loading completed**

Links to content: Adding article level metadata to WorldCat Local

	WorldCat Local Central Index Article Records	WorldCat Local Central Index Journal Titles Indexed
Year to Date Status (December)	440,692,000	68,946

Clustering and Display of Article Records

- Enhanced clustering
 - Working to improve clustering of records representing the same article this Fiscal Year
 - Improve user experience by grouping these records together while maintaining the ability for users to access the articles to which they have access.
 - Significant effort based on complex data analysis and development.

Better links: brought to you by the WorldCat knowledge base

Included in a standard cataloging subscription

One place to manage a library's electronic holdings--both ejournals and ebooks--at the network level (i.e., in the cloud)

Collection or package level management of holdings

Controls display of journal/book/article level links in WorldCat Local

Enables resource sharing of e-articles (as licenses allow)

The WorldCat knowledge base

McGill University Library | Search | WorldCat | My Library Account (Sign in) | WorldCat (Sign in)

McGill
LIBRARY AND COLLECTIONS

nonlinear universe [Search](#)

Libraries to search: Libraries Worldwide | [Advanced Search](#)

[Print](#) [Share](#)

[Chat with a librarian](#)

Search results for 'nonlinear universe' limited to **Libraries Worldwide**

Narrow by Format

- ☒ All Formats (2,538)
- ☐ Internet resource (1434)
- ☐ Article (545)
- ☐ Chapter (33)
- ☐ Book (266)
- ☐ eBook (21)
- ☐ Thesis/dissertation (10)
- ☐ Computer file (262)
- ☐ VHS (8)

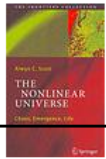
Refine Your Search


Author

- [Saha Bijan](#) (12)
- [Bertschinger Edmund](#) (6)
- [Conference Intern...](#) (5)
- [Hiroshima Univ](#) (5)
- [Silk J](#) (5)
- [Show more ...](#)

Results 1-10 of about 2,538 (.14 seconds)

[Select All](#) [Clear All](#) Save to: [New List] [Save](#) Sort by: Library and Relevance [Save Search](#)

1.  **[The nonlinear universe : chaos, emergence, life](#)**
by Alwyn Scott
eBook : Document
Language: English
Publisher: Berlin ; New York : Springer, ©2007
Database: WorldCat.org
Libraries that own this item: **McGill University Library**
[View Now](#) [View all editions and formats](#)

2.  **[The Nonlinear Universe: Chaos, Emergence, Life](#)**
by Alwyn C Scott, Robert E O'Malley Jr
Article : Peer-reviewed
Language: English
Publication: SIAM review. 50, no. 2, (2008): 401
Publisher: Philadelphia, Society for Industrial and Applied Mathematics.
Database: ArticleFirst
Libraries that own this item: **McGill University Library**
[View Now](#)

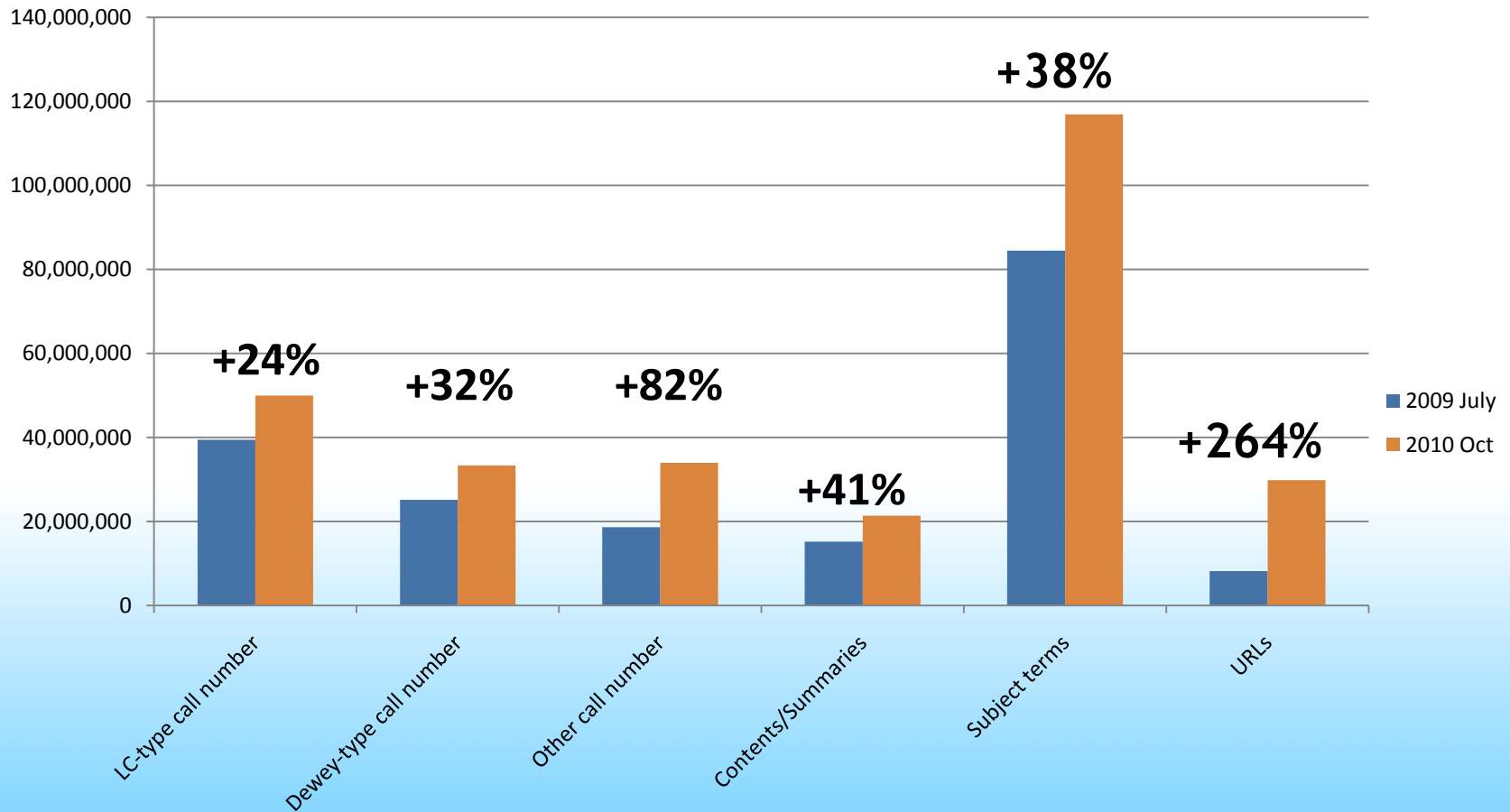
Both ebook and article level links appear together

The WorldCat knowledge base

- Future benefits:
 - Automatically set holdings in WorldCat
 - Automatically manage holdings by working with content providers resulting in improved quality of holdings
 - Integrated into acquisitions workflows
 - Display ebook links when a user discovers a record for the print book

ENRICHED RECORDS

Enriched information embedded in WorldCat records. Overall percent improvement 2009-2010: 49%



Enriched content from partners

- **Goal:** Offer a robust collection to support *discovery and selection by end-users*
- **Status:** 35 million data elements under contract
 - Book jacket covers, summaries, 1st chapters
 - Over 14.2 million ToCs
 - Over 360,000 music album covers (added Aug. 2010)
 - Over 1.5 million non-US book covers (added Nov. 2010)

Enriched content - looking ahead

- Add 207K more book covers
- Mine WorldCat itself to make the most of what we already have
 - Summaries AV and Books (9M)
 - Biographies (4.2M)

[<< Return to Search Results](#)



[Add to list](#)



[Add tags](#)



[Write a review](#)

Rate this item:



Hotel California

Author: [Eagles \(Musical group\)](#)

Publisher: [U.S.] : Asylum, [1999], 1976.

Edition/Format: Music CD : CD audio : Rock music : English [View all editions and formats](#)

Rating: (not yet rated) [0 with reviews - Be the first.](#)

Details - 'Hotel California'

Genre / Style: Rock; Country-Rock; Soft Rock; Contemporary Pop/Rock
Description: Country-Rock, Guitar (Electric), Vocals, United States of America, Soft Rock, Contemporary Pop/Rock, Album Rock
Release Date: Dec 8, 1976
Track Picks: [Track Picks](#): Hotel California; New Kid in Town; Life in the Fast Lane
Contents:

Title	Performer/Composer	Time
1 Hotel California	Felder/Frey/Henley	6:30
2 New Kid in Town	Frey/Henley/Souther	5:03
3 Life in the Fast Lane	Frey/Henley/Walsh	4:46
4 Wasted Time	Frey/Henley	4:56
5 Wasted Time (Reprise)	Frey/Henley/Norman	1:23
6 Victim of Love	Felder/Frey/Henley/	4:09
7 Pretty Maids All in a Row	Vitale/Walsh	3:58
8 Try and Love Again	Meisner	5:10
9 Last Resort	Frey/Henley	7:28

Reviews

[Editorial reviews](#)

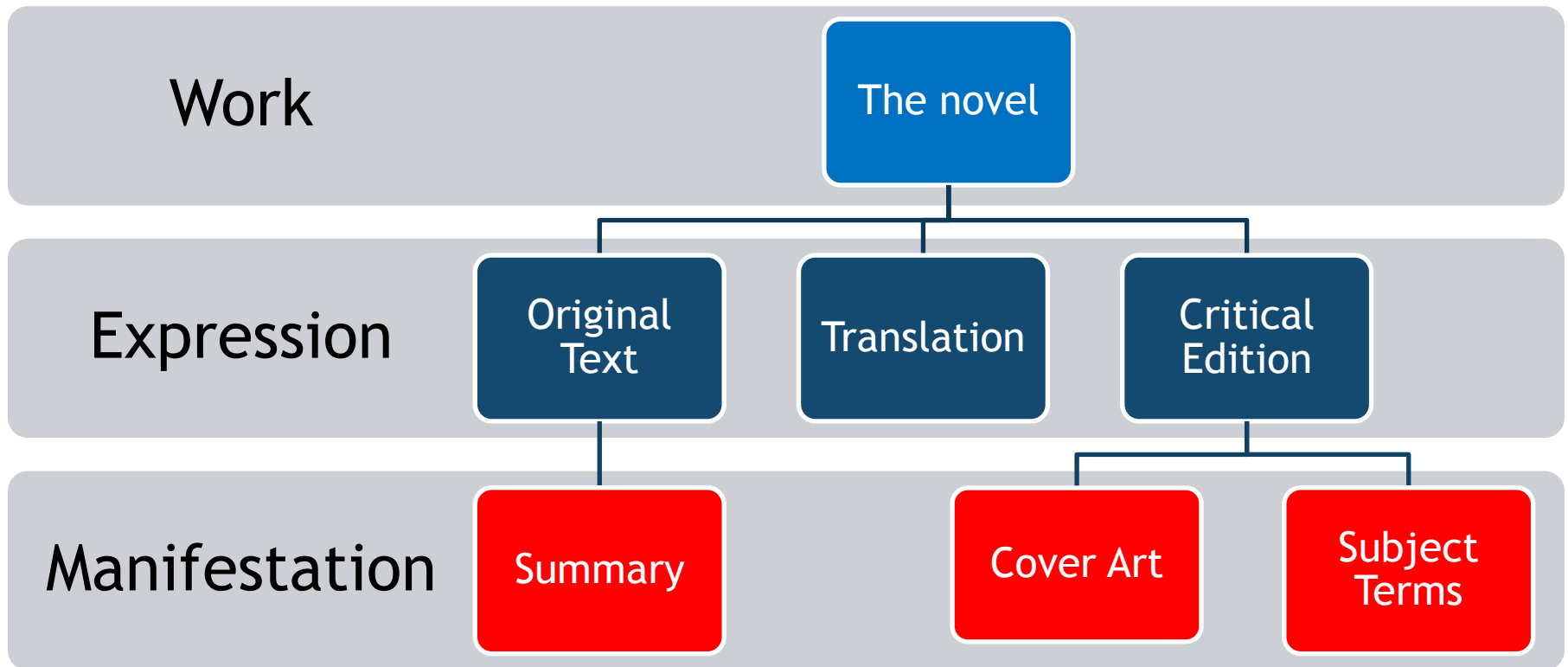
[All Music Guide](#)

Hotel California [Exclusive] [Track Picks](#)

The Eagles took 18 months between their fourth and fifth albums, reportedly spending eight months in the studio recording *Hotel California*. The album was also their first to be made without Bernie [Read more...](#)

Partner
Enrichment
Content

Mining WorldCat: Sharing data elements across a FRBR Work Set





Blade runner

25 edition(s), in English, held by 760 institution(s)

Summary

Los Angeles, 2019: Deckard, an expert "blade runner," must identify and execute four replicants, (genetically engineered beings virtually identical to humans) which have illegally returned to earth from their off-world slave duty.



[View All Summaries, Tables of Contents and Abstracts](#) (20)



[Find Any Edition In A Library](#)

EXPERIMENTAL

Details

Editions

25 edition(s), in 1 language(s), held by 760 institution(s)

Title/Author		Summary		Date	Lang	Format	Libraries
Blade runner Warner Bros. ; a Ladd Company production ; screenplay by Philip K. Dick ; produced by Michael Deeley ; screenplay by Michael Deeley ; directed by Ridley Scott. ⓘ		Los Angeles, 2019: Rick Deckard of the LAPD's Blade Runner unit prowls the steel & microchip jungle of the 21st century. His job is to track down and eliminate assumed humanoids known as 'replicants.' Replicants were declared illegal after a bloody mutiny on an Off-World Colony, and are to be terminated upon detection. He wants to get out of the force, but is drawn back in when 6 "skin jobs," the slang for replicants, hijack a ship back to Earth. The city that Deckard must search for his prey is a huge, sprawling, bleak vision of the future.		1991	English		223
Blade runner [a Ladd Company release in association with Warner Bros. ; presented by] Jerry Perenchio ; produced by Michael Deeley ; screenplay by Hampton Peoples ; directed by Ridley Scott. ⓘ		Los Angeles, 2019: Rick Deckard of the LAPD's Blade Runner unit prowls the steel & microchip jungle of the 21st century. His job is to track down and eliminate assumed humanoids known as 'replicants.' Replicants were declared illegal after a bloody mutiny on an Off-World Colony, and are to be terminated upon detection. He wants to get out of the force, but is drawn back in when 6 "skin jobs," the slang for replicants, hijack a ship back to Earth. The city that Deckard must search for his prey is a huge, sprawling, bleak vision of the future.		2007	English		180
Blade runner the Ladd Company in association with Warner Bros. ⓘ		Los Angeles, 2019: Rick Deckard of the LAPD's Blade Runner unit prowls the steel & microchip jungle of the 21st century. His job is to track down and eliminate assumed humanoids known as 'replicants.' Replicants were declared illegal after a bloody mutiny on an Off-World Colony, and are to be terminated upon detection. He wants to get out of the force, but is drawn back in when 6 "skin jobs," the slang for replicants, hijack a ship back to Earth. The city that Deckard must search for his prey is a huge, sprawling, bleak vision of the future.		1992	English		129
Blade runner the Ladd Company ; screenplay by David Peoples ; produced by Michael Deeley ; directed by Ridley Scott. ⓘ		Los Angeles, 2019: Rick Deckard of the LAPD's Blade Runner unit prowls the steel & microchip jungle of the 21st century. His job is to track down and eliminate assumed humanoids known as 'replicants.' Replicants were declared illegal after a bloody mutiny on an Off-World Colony, and are to be terminated upon detection. He wants to get out of the force, but is drawn back in when 6 "skin jobs," the slang for replicants, hijack a ship back to Earth. The city that Deckard must search for his prey is a huge, sprawling, bleak vision of the future.		2007	English		56
Blade runner Jerry Perenchio and Bud Yorkin present a Ridley Scott production ; screenplay by Hampton Peoples ; produced by Michael Deeley ; directed by Ridley Scott. ⓘ		Los Angeles, 2019: Rick Deckard of the LAPD's Blade Runner unit prowls the steel & microchip jungle of the 21st century. His job is to track down and eliminate assumed humanoids known as 'replicants.' Replicants were declared illegal after a bloody mutiny on an Off-World Colony, and are to be terminated upon detection. He wants to get out of the force, but is drawn back in when 6 "skin jobs," the slang for replicants, hijack a ship back to Earth. The city that Deckard must search for his prey is a huge, sprawling, bleak vision of the future.		2007	English		49
Blade runner a Ladd Company released in association with Warner Bros. ; Jerry Perenchio and Bud Yorkin present a Ridley Scott production ; screenplay by Hampton Peoples ; produced by Michael Deeley ; directed by Ridley Scott. ⓘ		Los Angeles, 2019: Rick Deckard of the LAPD's Blade Runner unit prowls the steel & microchip jungle of the 21st century. His job is to track down and eliminate assumed humanoids known as 'replicants.' Replicants were declared illegal after a bloody mutiny on an Off-World Colony, and are to be terminated upon detection. He wants to get out of the force, but is drawn back in when 6 "skin jobs," the slang for replicants, hijack a ship back to Earth. The city that Deckard must search for his prey is a huge, sprawling, bleak vision of the future.					

Download Complete Editions Table: XML CSV

Audience Level

MAKE IT EASIER TO CORRECT RECORDS; “SOCIAL CATALOGING”

Making it easier to correct the records: WorldCat community maintenance

Activity by Member Libraries during FY2010

	TOTAL
Expert Community	271,626
Database Enrichment	198,084
Minimal-Level Upgrade	176,618
Enhance Regular	176,491
Enhance National	45,451
CONSER Authentication	15,705
CONSER Maintenance	61,949
TOTAL	945,924

Expert Community

- Experiment conducted: mid-February 2009 through mid-August 2009
- All OCLC Cataloging members with full level authorizations were invited to participate – no application process
- Allowed member libraries with full-level Cataloging authos to make additions and changes to almost all fields in almost all records
- Great success - All functionality remains in place!

WORLDCAT STEWARDSHIP BY OCLC STAFF

OCCL Staff Maintenance Activity in FY 2010

	TOTAL
Bibliographic Records Replaced	12,511,044
Records Merged	150,992
Authority Records Created	1,977
Authority Records Replaced	94,744
CIP Records Upgraded	16,145

Other OCLC enhancements to WorldCat - a couple of recent examples

- Updated subject headings
 - Recent example: *Cookery* changed to *Cooking*
 - Over 314,000 records affected
 - ~75 new subject headings proposed to Library of Congress
- Adding Linking ISSNs (ISSN-L)
 - Added to about 800,000 records thus far

Additional OCLC staff enhancements

- Adding non-Latin cross-references to authority records
 - Almost 500,000 records affected
 - Non-Latin forms derived from WorldCat records
- Authority records for geographic names
 - Indirect subdivision forms added (about 90,000 records)
 - Geographic coordinates added to field 034 (more than 78,000 records)

OCCLC automated enhancements - looking ahead

- Looking ahead -
 - Automated heading control of name and subject headings (late FY11)
 - More automated enrichment of bibliographic records from mining FRBR work set data (late FY11)